优达学城数据分析师纳米学位

A/B 测试项目

一、试验设计

（一）指标选择

1、列出你将在项目中使用的不变指标和评估指标。

不变指标：选用 Number of cookies、Number of clicks 和 Click-through-probability

评估指标：选用 Gross Conversion 和 Net Conversion

2、对于每个指标，解释你为什么使用或不使用它作为不变指标或评估指标。此外，说明你期望从评估指标中获得什么样的试验结果。

Number of cookies：该数据产生于用户登录网站，不会因实验组中新增的页面而发生变化，因此选为不变指标。

Number of user-ids：该数据产生于试验发生之后，会受到试验的影响，所以它不能作为不变指标，但它可以作为评估指标。但是，由于实验组和对照组的 cookie 数量不一定完全相同，也就是说两组中用户 ID 数量的不同可能是由于实验的影响，也可能是由于两组 cookie 的不同造成的。所以使用用户 ID 数量的区别不能够很好的评估试验的效果。在一个比例化的评估指标（总转化率）存在的情况下，我们可以不选择用户 ID 的数量作为评估指标。综上，用户 ID 既不是不变指标，也不是评估指标。

Number of clicks：该数据产生于用户点击“免费试用课程”这个事件，不会因实验组中新增的页面而发生变化，因此选为不变指标。

Click-through-probability：该指标是 Number of clicks 与 Number of cookies 的比例，两个不变指标之商也必然是不变指标。

Gross conversion：为完成登录并报名参加免费试用的用户 ID 的数量与点击“开始免费试用”按钮的唯一 cookie 的数量之比。通过比例指标的运用，可以较好的弱化试验组与对照组之间因 cookies 数量不完全相同所造成的影响，因此选为评估指标。

Retention：是参加了“免费试用课程”14 天以后自动收费的用户 ID 数量与注册参加“免费试用课程”的用户 ID 数量的比例。但是，本实验的 unit of diversion 是 cookies，而 Retention 的 unit of analysis 是 user-id，unit of diversion 与 unit of analysis 不相同，无法确保引流的一致性，从而造成分析变异性与经验变异性不匹配，若时间允许，该指标可采用经验方法计算，本文拟直接舍弃。

Net conversion：是参加了“免费试用课程”14 天以后自动收费的用户 ID 数量与点击“开始免费试用”按钮的唯一 cookie 的数量之比。该指标同样受到试验影响且为比例指标，因此选为评估指标。

期望从评估指标中获得的试验结果：

若试验组的 Gross conversion 显著低于控制组，则表明新增的页面能够有效阻止学习时间不足的用户进行注册；

若试验组的 Net conversion 不显著低于控制组，则表明新增的页面使付费用户减少的程度不大。

（二）测量标准偏差

1、列出你的每个评估指标的标准偏差。

Gross conversion 的 Standard deviation 为 0.0202。

Net conversion 的 Standard deviation 为 0.0156。

2、对于每个评估指标，说明你是否认为分析估计与经验变异是类似还是不同（如果不同，在时间允许的情况下将有必要进行经验估计）。简要说明每个情况的理由。

本实验选择的分组单元是 cookie，评估指标 Gross conversion 及 Net conversion 的分析单元也均为 cookie，可以确保引流的一致性，所以这两个指标的分析差异性匹配经验差异性。

（三）规模

1、样本数量和功效，说明你是否会在分析阶段使用 Bonferroni 校正，并给出实验正确设计所需的页面浏览量。

不会在分析阶段使用 Bonferroni 校正，因为 Gross conversion 和 Net conversion 具有关联性，使用 Bonferroni 校正得到的结果过于保守。

使用 alpha=0.05，（1-beta）=0.8。

Gross conversion：dmin=0.01，Baseline conversion rate = 20.625%，根据在线计算器得到一组实验所需 Gross conversion 的样本量为 25835，两组实验所需的样本量为 51676，然后转化为需要的页面浏览量 = 51676/0.08 = 645950

Net conversion：dmin = 0.075，Baseline conversion rate = 10.93125%，根据在线计算器得到一组实验所需 Net conversion 的样本量为 27413，两组实验所需的样本量为 54826，然后

转化为需要的页面浏览量 = 54826/0.08 = 685325

考虑到所需页面浏览量要同时覆盖两个指标，因此选择两个页面浏览量中较大的一个，即 685325。

2、持续时间和暴光比例。说明你会将多少百分比的页面流量转入此试验，以及鉴于此条件，你需要多少天来运行试验。

一般来说，A/B testing 的实验时间是持续几个星期至一个月之内。根据所需的页面浏览量（685325）及每天的页面浏览量（40000），曝光的流量部分为 57.2%时所需天数为 30 天，符合要求。因此选择曝光比例为 57.2%，试验持续时间为 30 天。

3、说明你选择所转移流量部分的原因。你认为此试验对优达学城来说有多大风险？

不选择对所有流量开展实验的原因主要是出于以下几个方面：

一是安全性。推出这个新的页面弹窗，我们不确定它是否能在所有浏览器中正常运行，也不确定用户将有什么反应，所以选择仅向部分用户开展实验。

二是随机分配分组单元时，为了避免异常数据（例如节假日等等）对试验结果的误导，也倾向于压缩发送流量比例，在合理范围内延长持续时间，从而尽可能的了解用户在不同日期的差异。

对此实验，优达学城并没有太大的风险，因为：

第一，即使学生每周学不到五小时，他们只是被页面的变更提醒引导到了另外的一个页

面，如果今后有需要学生仍然可以进入免费试学、登陆并可能完成课程，不会因此影响用户使用网站的习惯；

第二，没有在页面展示上有过大的改动，不会对用户产生感情上的冲击，用户也不需要花长时间去适应页面的改变。

第三，该试验没有关于数据库及后台的改变，不用担心数据的丢失及由于后台的失误导致网页奔溃用户无法访问网页等大问题。

第四，此试验也不会对用户的个人信息安全造成风险，因为不论网页是否增加了提醒，用户在确认参加免费试学时都需要输入信用卡信息，而很明显系统一定会保护用户的个人信息。

第五，该试验同样也没有道德上的风险。

二、试验分析

（一）合理性检查。对于每个不变指标，对你在 95%置信区间下期望观察到的值、实际观察的值及指标是否通过合理性检查给出结论。

Number of cookies：控制组总计有 345543 个观测值，试验组总计有 344660 个观测值。对于每一个观测值，它被发送至控制组和试验组的几率均为 50%，且事件之间相互独立，因此符合二项分布特征。在 alpha = 0.05 水平下，SE = ((0.5*0.5) / (N-con + N-exp))^0.5 = 0.0006，margin of error = SE * Z-score = 0.0006 * 1.96 = 0.0012，围绕 0.5 为中心的置信区间为(0.4988,0.5012)。实际观测值= N-con / (N-con + N-exp) = 0.5006，位于置信区间之内，因此 Number of cookies 通过 Sanity check。

Number of clicks：控制组总计有 28378 个观测值，试验组总计有 28325 个观测值。对于

每一个 click 事件，它被发送至控制组和试验组的几率均为 50%，且事件之间相互独立，因此符合二项分布特征。在 alpha = 0.05 水平下，SE = ((0.5*0.5) / (N-con + N-exp))^0.5 = 0.0021，margin of error = SE * Z-score = 0.0021 * 1.96 = 0.0041，围绕 0.5 为中心的置信区间为(0.4959,0.5041)。实际观测值= N-con / (N-con + N-exp) = 0.5005，位于置信区间之内，因此 Number of clicks 通过 Sanity check。

Click-through-probability：控制组的 P-con = X-con / N-con = 0.0821，试验组的 P-exp = X-exp / N-exp = 0.0822。试验组与控制组的差异：P-exp – P-con = 0.0822 – 0.0821 = 0.0001。两组的合并概率 P-pool = (X-con + X-exp) / (N-con + N-exp) = 0.0822，合并标准误差 SE-pool = (P-pool * (1 – P-pool) * (1/N-con + 1/N-exp))^0.5 = 0.0007，则 Margin of error = SE-pool * Z-score = 0.0007 * 1.96 =0.0013，围绕 0 为中心的置信区间为(-0.0013, 0.0013)。两组的差异 0.0001 位于置信区间之内，因此 Click-through-probability 通过 Sanity check。

所有不变指标均通过 Sanity check。

（二）结果分析

1、效应大小检验。对于每个评估指标，对试验和对照组之间的差异给出 95% 置信区间。说明每个指标是否具有统计和实际显著性。

Gross conversion：

P-pool = (X-con + X-exp) / (N-con + N-exp) = (3785+3423) / (17293+17260) = 0.2086

SE-pool = (P-pool*(1-P-pool)*(1/N-con + 1/N-exp))^0.5 = 0.0044

Margin of error = SE-pool * Z-score = 0.0086

d-hat = P-exp – P-con = X-exp/N-exp – X-con/N-con = -0.0206

CI: (-0.0291 , -0.0120)

置信区间不包括 0，具有统计显著性，不包含最小实质显著性边界-0.01，具有实质显著性。

Net conversion:

P-pool = (X-con + X-exp) / (N-con + N-exp) = (2033+1945) / (17293+17260) = 0.1151

SE-pool = (P-pool*(1-P-pool)*(1/N-con + 1/N-exp))^0.5 = 0.0034

Margin of error = SE-pool * Z-score = 0.0067

d-hat = P-exp – P-con = X-exp/N-exp – X-con/N-con = -0.0049

CI: (-0.0116 , 0.0019)

置信区间包括 0，不具有统计显著性，包含最小实质显著性边界-0.0075，不具有实质显著性。

2、符号检验。对于每个评估指标，使用每日数据进行符号检验，然后报告符号检验的 p 值以及结果是否具有统计显著性。

Gross conversion:

将每日试验组与控制组的 Gross conversion 做差，为正值的有 4 天，全部观测天数为 23 天，使用在线计算器得到双尾 p 值为 0.0026，小于 alpha = 0.05，因此具有统计显著性。

Net conversion:

将每日试验组与控制组的 Net conversion 做差，为正值的有 10 天，全部观测天数为 23

天，使用在线计算器得到双尾 p 值为 0.6776，远大于 alpha = 0.05，因此不具有统计显著性。

3、汇总。说明你是否使用了 Bonferroni 校正，并解释原因。若效应大小假设检验和符号检验之间存在任何差异，描述差异并说明你认为导致差异的原因是什么。

未使用 Bonferroni 校正。因为 Gross conversion 与 Net conversion 具有相关性，而 Bonferroni 校正在此种情况下过于保守。此外，如果要进行 Bonferroni 校正的话，除了判断指标是否相互独立，我们还要判断几个评估指标之间是“与”还是“或”的关系。如果指标之间是相互独立的，但是得出最终结论是“与”的关系，这样我们进行 Bonferroni 校正也会过于保守。在本试验中我们希望这两个试验结果同时满足，因此也不需要使用 Bonferroni 校正。

效应大小假设检验与符号检验之间未存在差异。

三、建议

Gross conversion 具有统计和实际显著性，是我们希望看到的结果。但是 Net conversion 的置信区间包含负数，根据此处的计算结果(-0.0116, 0.0019)，也就是说有很大的概率 Net conversion 会减少，并且有一定的概率 Net conversion 的减少会超过实际显著性 0.0075。因此我们无法说明”降低的程度不大“。所以不建议发布该变更。

四、后续试验

对你会开展的后续试验进行概括说明，你的假设会是什么，你将测量哪些指标，你的转移单位将是什么，以及做出这些选择的理由。

在此试验的基础上，对于那些参加了 14 天免费课程却点击取消的用户增加弹窗，询问是否愿意接受更为专业的教练辅导和项目审阅服务，并在弹窗中附往期毕业学员成功就职名企的简述和详细页面链接。若点击是，提示学员将在 14 天免费试用到期后自动扣款，若点击否，则关闭弹窗确认取消课程。

假设：参加免费课程的学员有可能因为课程内容难度过大而放弃，也有可能因为学习课程后能否顺利就业等存有疑虑。通过提示用户付费后能够接受更为专业的辅导和服务及往期毕业学员成功就职名企的信息，能够有效打消拟取消课程用户的疑虑，增加试用用户付费比例。

测量指标包括：

1. 参加 14 天免费课程的 user-ids 数量
2. 14 天免费课程结束后自动扣费的 user-ids 数量
3. 付费率：即 14 天免费课程结束后自动扣费的 user-ids 数量占参加 14 天免费课程的 user-ids 数量的比例

转移单位：由于以上测量指标全部为 user-ids 的数量，所以转移单位自然选择 user-ids 的数量。